
Microbiological Source Tracking Workshop: Summary of Proceedings

**Brock B. Bernstein
John F. Griffith
Stephen B. Weisberg**

March 2002

Introduction

Microbiological source tracking methods are potentially powerful tools that are increasingly being used to define the nature of water-quality problems in watersheds across the nation. While these techniques show much promise, most are still in the early stages of development. Many of these techniques have been tested in a limited number of watersheds and with a limited number of possible sources.

Public agencies, particularly those in California, are preparing to spend several million dollars in applying these techniques to identify sources of bacteriological water contamination. These expenditures are based upon the assumption that these methods provide a firm scientific foundation for regulatory or remediation decisions.

This workshop brought together nationally recognized experts in environmental microbiology, molecular biology, and microbial detection methods with the intention of better defining the current state of knowledge regarding source tracking techniques, standardizing field and laboratory methods for those approaches that are most well developed, and drafting a protocol for a National Source Tracking methods comparison study that can be used to inform local decision makers about the reliability of the different methods.

Sponsored by the U.S. Environmental Protection Agency, California State Water Resources Control Board, Southern California Coastal Water Research Project, and National Water Research Institute, this workshop represented a unique opportunity for cooperative interaction among those involved in all facets of microbiological water quality.

Workshop Structure

The Microbial Source Tracking Workshop was held February 5 – 7, 2002 in Irvine, California. The main intent of the Workshop Steering Committee (Appendix 1) was to summarize existing knowledge about source tracking methods and to then use this review as the basis for designing studies to compare, evaluate, and validate a wide range of such methods. The intended product of the workshop was the outline of an approach for carrying out such studies.

The first day of the workshop was a public session attended by 226 people. This session was comprised of 15 presentations on source tracking issues and methods, including ribotyping and other genetic fingerprinting approaches, antibiotic resistance profiles, biomarkers, and more specific indicator organisms (Appendix 2). The variety of methods presented can be broadly categorized according to two key sets of features (Table 1). First, there are methods that distinguish among bacterial and/or viral samples based directly on their genetic makeup (genotypic methods) and those that distinguish among samples based on secondary characteristics such as antibiotic resistance (phenotypic methods). Second, there are methods that require the development of a background library or database against which to compare a sample and those that do not need such a library for their completion. The presentations focused on recent testing or applications of these methodologies. Many presenters indicated that they are pursuing multiple techniques.

Table 1. Two-way classification of some of the more widely used source tracking methods in terms of their focus on genotypes or phenotypic characteristics and their relative dependence on a background library or database of genotypic or phenotypic characteristics.

	<i>Library dependent</i>	<i>Library independent</i>
<i>Genotypic</i>	ribotyping bacterial community fingerprinting (t-RFLP) repetitive intergenic DNA sequences (PCR)	F+ coliphage human pathogenic virus detection (PCR or RT-PCR) Bacteroides genotyping (PCR) Enterotoxin biomarkers (PCR)
<i>Phenotypic</i>	multiple antibiotic resistance carbon source profiling	Enterotoxin biomarkers F+ coliphage serotyping IgA Antibodies

The second and third days of the workshop involved small-group discussion sessions that were limited to 40 invited participants. There were three such sessions, which concentrated on: 1) identifying a set of criteria to use in evaluating the outcomes of methods comparison studies, 2) design of such studies, and 3) defining adequate library size for the library-dependent methods. The invited participants were divided into four breakout groups of 10 people each according to affiliation and expertise. All groups were asked to address the same issues. After each session, all 40 participants were reconvened to compare recommendations among groups. The following sections of this report summarize the recommendations that evolved from the small group sessions.

Methods Evaluation Criteria

Four subgroups of workshop participants discussed conceptual, technical, policy, logistical, and economic issues involved in improving the application of microbiological source tracking methods to the point where they could be used consistently, broadly, and with confidence in a range of real-world conditions. There was unanimous agreement that a rigorous methods comparison and standards development effort is needed because the current situation is characterized by a lack of uniform standards or reference materials and ad hoc rules of thumb for sampling and measurement designs. After reviewing the work of the four separate subgroups, the participants agreed that the following set of criteria should provide the basis for any methods comparison study. These are divided into three tiers (measurement reliability, management relevance, cost and logistics) that reflect distinct categories of issues identified in the discussion (Table 2).

The ordering of the tiers reflects an inherent prioritization of the evaluation criteria. The most fundamental requirement of any method is that it provides accurate and repeatable results (Tier 1). Without this, other criteria, such as ease of communication to the public or a low cost per sample, would be meaningless. Likewise, adequate performance on the management criteria (Tier 2) is more important than cost and logistical issues (Tier 3). For example, a small library size or minimal training requirement would be less important if test results were not related to public health outcomes of interest or could not be readily communicated to key management audiences.

Table 2. Method evaluation criteria agreed on by the workshop participants. Criteria are divided into three categories, or tiers, that reflect different aspects of performance.

Category of criteria	Specific evaluation criteria
Tier 1: Measurement reliability	<ul style="list-style-type: none"> • Reproducibility of results within and across laboratories • Accuracy of classification of isolates into the correct group of sources (for library dependent methods) • Confidence that an identified indicator is from the presumed source (for library independent methods) • Level of resolution, or ability to discriminate among sources (i.e., human vs. non-human, livestock vs. wildlife, non-human species level, cattle from separate farms) • Matrix stability (in what matrices, e.g., saltwater, freshwater, turbid water, humic acid environments, is the method applicable?) • Geographical stability (over what area is the method applicable?) • Temporal stability (over what time frame is the method applicable?) • Confirmation by peer review
Tier 2: Management relevance	<ul style="list-style-type: none"> • Relationship to actual source(s) of contamination • Relationship to public health outcomes • Relationship to commonly used water quality indicators • Ease of communication to the public • Ease of communication to management audiences
Tier 3: Cost and logistics	<ul style="list-style-type: none"> • Equipment and laboratory facilities required • Training required • Library size required (for library dependent methods) • Library development effort per "unit" required (for library dependent methods) • Implementation time • Cost of ensuring results are legally defensible • Cost per sample, including all operations and maintenance overhead • Sample turnaround time

Tier 1 focuses on measurement issues that affect the accuracy and repeatability of test results. While accuracy and repeatability are important for any analytical method, there are several aspects of this that are specific to microbiological source tracking methods. For all such methods, the lack of widely accepted standardized techniques makes the issue of reproducibility of results both within and across laboratories more crucial than it might be for methods with a longer history of development and application. Another measurement issue involves how accurately different methods identify the source(s) of a particular isolate or indicator from a sample. While the manner in which this issue would be evaluated differs somewhat depending on whether the method depends on a library of isolates, the core question of accurate identification of isolates and/or indicators is basic to the acceptance and wider use of any method.

The working groups identified additional Tier 1 criteria related to a set of "stability" issues that are associated with repeatability. These issues include geographic, temporal, and matrix stability. Geographical stability refers to the size of the area over which a particular method (and the library or indicators it is dependent on) is applicable. Species of gut flora in a particular host species may vary spatially for a variety of reasons, and even the same bacterial or virus species

may vary genetically from one area to another, sometimes significantly so. Temporal stability refers to a method's applicability over time. The gut flora community of a host species, or even a host individual, has the potential to shift over time. Methods or indicator species that are robust to such changes are preferable to methods which require more frequent library renewal. Matrix stability refers to a method's performance in the presence of interferences that may obstruct or bias the method. For example, saltwater can inhibit culturing of some organisms. Organic compounds, such as humic acid, can impede PCR because it binds to DNA present in the sample, making the DNA unavailable to the enzymes used in the method. Depending on the particular kind of interference, methods will vary in their sensitivity to these factors.

Tier 2 focuses on management issues that affect the acceptance and usefulness of test results in decision making. The more directly a method targets actual sources of contamination, such as differentiating among all animal types as compared to only differentiating human from non-human sources, the more valuable it will be for decision making. A particular method is similarly strengthened the more it can be linked, through epidemiological studies or other means, to public health outcomes of concern. On the other hand, workshop participants understood that currently used water quality indicators, despite their weaknesses in these two respects, are widely used, have a long time series of data, and form the basis for regulatory policy. Microbiological source tracking methods that have a demonstrated relationship to the currently used water quality indicators would be preferable. Finally, preferable methods should readily understood by both public and management audiences. Ease of communication will help promote their widespread acceptance and use.

Tier 3 focuses on cost and logistical issues directly related to implementation. These are relatively straightforward and include items such as the expense of necessary measurement equipment and the degree of staff training required.

Recommended Study Design

The participants at the workshop unanimously agreed on the need for rigorous studies to compare methods on the basis of the criteria described above. The participants agreed on a study design structured around four phases that advance through successively more complex questions about method performance, beginning with single-source, within-laboratory studies and ending with diverse, multi-source watershed evaluations. The cost efficiency of the study could be maximized if the sources and/or locations involved in the early phases were continued through the later phases. The following sections describe each of the four phases:

Phase 1: Repeatability

The core question to be addressed in Phase 1 is: How repeatable is the measurement technique? More specifically, the participants wanted to determine if the a laboratory can produce the same results for the same samples in multiple runs of the same method. Moreover, they thought it was important to find out if multiple investigators in different laboratories could produce the same results for the same samples.

In this initial phase, multiple replicate samples from just a few unique sources would be distributed to a series of different laboratories to assess repeatability of each method both within- and between-laboratories. Some of the samples would be distributed as "knowns" associated

with a specific source, while others would be distributed as “blind” samples to assess the investigator’s ability to match patterns with the known samples. In either case, samples in this phase would be distributed as isolates or in the simplest, most replicable format associated with that particular method. Since one purpose of this phase is to quantify repeatability among laboratories using the same method, considerable method standardization would be required prior to testing.

Several of the breakout groups suggested that this phase could be extended to also address temporal stability. Two types of temporal stability testing were identified. First, selected samples could be retained in the laboratory and processed at several successive time intervals. Second, fecal material from the same animal could be sampled over time to assess how repeatable or applicable a library developed at one time is to an environmental sample collected at a later time.

Phase 2: Accuracy – Laboratory samples

The core question to be addressed in Phase 2 is: Can methods accurately detect the source(s) in laboratory-created “blind” samples? This phase would be the next step in a continuum of increasingly sample complexity and would differ from Phase 1 in three ways. First, test samples would be distributed in an aqueous matrix to simulate the manner in which environmental samples would be received in a typical application. Second, a subset of the samples would include feces from multiple sources, again to more closely simulate the conditions encountered in typical applications. Third, the tests could be extended to incorporate matrix interferences (e.g. turbidity, salt) into the test samples.

This phase would proceed in two steps. In the first step, participating laboratories would be provided with the raw fecal material and/or isolates (from individual sources) required to build libraries needed to carry out the tests. Participating laboratories would be told the sources of these samples and participants generally agreed that laboratories should be allowed to develop libraries of whatever size they deemed appropriate. Participants also agreed that these samples should include the sources used in Phase 1, as well as those likely to be used in Phases 3 and 4, in order to promote efficiency by reducing the amount of repetitive library development that would otherwise be required through all four study phases. In addition, the sources used in Phases 1 and 2 should reflect sources that are of primary concern in watershed characterization studies and TMDL development. This will not only include an important aspect of realism in the study, but will also help promote interest in, and support for, the further application of microbial source tracking methods.

In the second step of Phase 2, participating laboratories would be provided a series of “unknowns” to identify. These samples would represent both single sources and various mixtures of sources created in the laboratory and placed into a water sample. For example, if four sources are used, laboratories might receive samples comprised of single sources and combinations of two, three, and all four sources. In addition, these samples could also be provided in different matrices, with the most likely ones mentioned by workshop participants being salt water, fresh water, turbid water and water containing humic acid. This would be intended to assess robustness of the methods when potential interferences are present.

Several participants expressed concern about the difficulty of combining different sources into a common sample, given the different pathogen density per unit fecal material among source types, but several other participants suggested that this could be accomplished by accurate quantification

of indicator bacteria density in the source material. Other participants suggested that requiring all laboratories to process the full set of samples (i.e., multiple combinations of sources in all matrices) would involve a large effort for each laboratory, which would be exacerbated if individual laboratories were to test more than one method. While no consensus was reached on the issue of laboratory workload, options were identified for randomizing the assignment of samples to laboratories that would allow for individual laboratories to test only a portion of the full complement of potential combinations of sources, matrices, and methods.

Phase 3: Accuracy – Ambient samples

The core question to be addressed in Phase 3 is: Can methods accurately detect source(s) in simple, dominant-source watersheds?

Phase 3 would extend the study to the testing of ambient samples (e.g., of water or soil) collected directly from the environment. However, in order to fit with the overall study strategy of proceeding in incremental steps of increasing complexity, the workshop participants agreed these samples would be drawn from a relatively simple watershed dominated by one or only a few sources. While this watershed might include the same sources used in Phase 2, samples could be drawn, for example, from runoff directly below known sources (i.e., dairy farm, leaking septic system) and would thus constitute a performance test of methods in a real-world situation, albeit a relatively simplified one. This approach will extend the Phase 2 tests by including soil bacteria, potential mixtures of sources, and other real-world interferences that cannot be accurately mimicked in laboratory-created samples. Thus, while the dominant source in the test samples will be known to the test managers, the actual makeup of the samples will be unknown both to the test managers and to the laboratories themselves. Depending on the results of Phases 1 and 2, test managers could calibrate the complexity and difficulty of Phase 3 samples by selecting test samples at different distances downstream of major sources within the watershed. This assumes that test samples collected further downstream will include more potential interferences.

Phase 3 will require careful selection of the test watershed. Workshop participants emphasized using a single source catchment. Some participants felt that having characterization data available on the test watershed might be valuable. Such information might include source identification from sanitary surveys, census data, wildlife surveys, and other studies, and should include detailed GIS coverages that enable mapping of test results if needed. This could enable test managers to develop a more rigorous set of expectations against which to evaluate the performance of participating laboratories.

Once again, several workshop participants recognized that using the same sources identified in Phase 2 will increase cost-efficiency, particularly for the library-dependent methods, and will ensure that geographic variability is minimized at the same time. Other participants suggested that an assessment of the geographic stability of methods could be incorporated into Phase 3 by expanding the design to include additional populations and/or watersheds.

Phase 4: Comparability in complex watersheds

The core question to be addressed in Phase 4 is: Do different methods produce comparable results in more complex systems with unknown sources?

Phase 4 would essentially mimic Phase 3, but with samples drawn from more complex watersheds with a wider variety of sources. While Phase 4 assesses whether different methods provide the same answer, it does not allow for easy explanation of differences among methods because underlying source materials and potential interferences are unknown. The workshop participants devoted somewhat less attention to this phase of the study, under the assumption that more detailed design decisions would be dependent on the results of the first three phases. However, they did stress that it should include, as a subset, the sources and watershed(s) used in Phases 1 – 3.

Depending on the number of sites, watersheds, and/or populations used, Phase 4 will permit a more thorough assessment of the geographic stability criterion as well as a start toward an assessment of the temporal stability criterion under the Tier 1 set of evaluation criteria, depending on the length of time the evaluation study continues.

Evaluating management, cost, and logistical criteria

Data from Phases 3 and 4 will also permit an evaluation of the Tier 2 set of evaluation criteria related to management, particularly the relationship of test results to actual sources of contamination and commonly used water quality indicators. Evaluation of the relationship of test results to public health outcomes will depend on the availability of adequate epidemiological data.

Each of the four phases will provide quantitative information needed to evaluate the cost and logistical criteria. This information can be combined with the broader experience of practitioners to develop a comprehensive assessment of the cost and logistical constraints and implications of each method.

Two of the breakout groups identified that part of evaluating cost-efficiency of a method involved quantifying multiple sources of variability inherent in applying the method. Methods with smaller cumulative variability within a source type will be more cost-effective, as they will require less, or less frequent, library development. Additionally, understanding these sources of variability provides the basis for developing a formal statistical model and would help furnish a basis for decision making about specific study design issues such as the necessary number of replicate samples. Sources of variability were grouped into four major categories:

- Laboratory
 - Within laboratory
 - Across laboratories
- Space
 - Within isolate
 - Across isolates within animal or sample
 - Across animals within herd or restricted population
 - Across herds or restricted populations of the same species within a watershed
 - Across watersheds
 - Across larger systems
- Time
 - Short-term
 - Seasonal
 - Year-to-year

- Matrix
 - Freshwater
 - Saltwater
 - Turbid water
 - Humic environments

Library Development

Practitioners have been making decisions about library size for the library-dependent methods based on logistical and cost constraints and/or on ad hoc rules of thumb developed through practical experience. No studies, however, have yet rigorously evaluated the minimum or optimal library size in terms of the kinds of evaluation criteria listed above. The workshop participants agreed unanimously on the need for a thorough study of this issue.

Library size is directly related to the accuracy of classification, matrix stability, level of resolution, and geographic and temporal stability criteria under the Tier 1 set of evaluation criteria. Participants agreed that this is an optimization problem in which the goal is to find a point of diminishing returns after which extra sampling to build a larger library adds less and less to a method's performance on the relevant evaluation criteria. A key question for both managers and practitioners is the degree to which a library developed for a particular watershed or region will apply to other watersheds or regions. A useful approach to this question recommended during the workshop was to define a set of nested watersheds, selected in coordination with Phases 3 and 4 of the study design described above, and determine the optimal library size for successively larger geographic areas.

The workshop participants also reached several broad conclusions about the issue of library development. First, they agreed that there is not enough knowledge at present to determine optimal library size for the library dependent methods. Second, they agreed that the optimal size would, to some extent, probably be method and watershed specific. Given the way that libraries are used in different methods, as well as differences in genetic variability across target organisms, it is highly unlikely that a single size library will prove optimal in all situations. Third, they agreed that statistical analyses of the large amount of existing data, if properly integrated, could provide useful answers about the sources of variability listed above and therefore about optimal library size. Fourth, they agreed that such analyses would not replace the need for a rigorously designed study that would specifically address the issue of library size. The ultimate goal of such efforts would be to develop an ability to model or predict the size of a needed library based on watershed characteristics such as the identity of sources and the loading processes that move contaminants through a system.

Thus, near-term decisions about library size will necessarily have to be based on best judgment and practical experience. There was general support among workshop participants for the idea that, in the near future, it would be possible to develop some preliminary guidelines based on statistical analyses of existing data. SCCWRP volunteered to begin such analyses as soon as researchers provide data. Two analyses discussed included a preliminary assessment of geographic variability in isolates and an analysis of the breakeven point at which the number of unique patterns identified per number of isolates begins to level off. Over the longer term, the results of a more rigorous evaluation would improve on these preliminary guidelines. Finally, while participants believed that developing a national library of geographically stable isolates would be of benefit, they agreed that more local and regional work should be carried out first.

Conclusion

Workshop participants were enthusiastic about the rapid advances in source tracking techniques that had been made in the last several years. They also felt that many of the techniques presented at the workshop would fare well in the tests described above and such tests would dramatically improve the rigor of microbial source tracking methods. There was recognition, though, that because of the diversity of animal types, as well as intrinsic evolutionary processes in microbial organisms, these methods will never be 100% accurate. The challenge for practitioners is to recognize that there will always be some patterns that cannot be classified and to determine the point at which increased effort devoted to refining methods and building libraries begins to produce diminishing returns.

The workshop participants were also unanimous in agreeing on the value of the workshop. Recognizing that the field will continue to advance rapidly over the next several years, participant agreed on the need for additional periodic workshops in the future, perhaps every one or two years.